

## **PENERAPAN DATA MINING DENGAN ALGORITMA ID3 UNTUK MEMPREDIKSI PENJUALAN (STUDI KASUS : PT. TATA WARNA CIPTA PERKASA)**

\* Jepri Banjarnahor, Erick Reinaldo, Evta Indra  
Program Studi Sistem Informasi Fakultas Teknologi dan Ilmu Komputer Universitas Prima Indonesia  
E-mail : \* jepribanjarnahor@unprimdn.ac.id

**ABSTRAK-** Di zaman teknologi ini, masih banyak pemanfaatan teknologi yang tidak maksimal, misalnya pada bidang pengolahan data, banyak perusahaan besar yang memiliki permasalahan pada bidang ini. PT. Tata Warna Cipta Perkasa merupakan perusahaan berada di Pematangsiantar yang bergerak di bidang penjualan bahan bangunan berupa baja ringan seperti rangka baja ringan, reng, dan gypsum. PT. Tata Warna Cipta Perkasa memiliki permasalahan pada pengolahan data stock barang. Permasalahan ini mengakibatkan penumpukan stock barang, tidak terpenuhinya kebutuhan konsumen, dan tidak adanya data yang jelas tentang stock barang.

**Kata kunci :** Data Mining, Algoritma ID3, Decision Tree

### **1. PENDAHULUAN**

Di zaman teknologi ini, masih banyak pemanfaatan teknologi yang tidak maksimal, misalnya pada bidang pengolahan data, banyak perusahaan besar yang memiliki permasalahan pada bidang ini.

PT. Tata Warna Cipta Perkasa merupakan perusahaan berada di Kota Medan yang bergerak di bidang penjualan bahan bangunan berupa baja ringan seperti rangka baja ringan, reng, dan gypsum.

Permasalahan pada pengolahan data stock barang. Permasalahan ini mengakibatkan penumpukan stock barang, tidak terpenuhinya kebutuhan konsumen, dan tidak adanya data yang jelas tentang stock barang.

Untuk mengatasi permasalahan tersebut di atas, maka dilakukan prediksi penjualan pada PT. Tata Warna Cipta Perkasa menggunakan penerapan data mining dengan metode pohon keputusan algoritma ID3.

Penelitian ini dilakukan dengan mengolah data penjualan bahan bangunan berupa baja ringan dan kemudian mengklasifikasikan penjualan bahan bangunan tersebut. Untuk pengolahan data dan informasi digunakan metode klasifikasi data mining yakni menggunakan metode pohon keputusan menggunakan algoritma ID3.

### **STUDI PUSTAKA**

#### **A. Pengertian Data Mining**

Data mining juga disebut sebagai serangkaian proses untuk menggali nilai tambah berupa pengetahuan yang selama ini tidak diketahui secara manual dari suatu kumpulan data (Pramudiono, 2006 / Kusriani dan Luthfi, 2009).

#### **B. Tahapan Proses KDD**

Pada bagian ini akan dijelaskan proses data mining yang terdiri dari (Kusnawi, 2007):

1. Pembersihan data (data cleaning), untuk membersihkan noise dan data yang tidak

2. konsisten. Dalam kasus ini membersihkan data-data pasien yang sudah dihapus dan identitas yang tidak lengkap (misal: umur, status marital, pendidikan, diagnosa, dan sebagainya);
3. Integrasi data, penggabungan data dari berbagai sumber;
4. Transformasi data, data diubah menjadi bentuk yang sesuai untuk di mining;
5. Aplikasi teknik data mining, proses inti dimana teknik data mining diterapkan untuk mengekstrak pola-pola tertentu pada data;
6. Evaluasi pola yang ditemukan;
7. Presentasi pengetahuan, menggunakan teknik visualisasi untuk menampilkan hasil data mining kepada pengguna user

#### **C. Decision Tree**

Dalam decision tree tidak menggunakan vector jarak untuk mengklasifikasikan obyek. Seringkali data observasi mempunyai atribut-atribut yang bernilai nominal, misalkan obyeknya adalah sekumpulan buah-buahan yang bisa dibedakan berdasarkan atribut bentuk, warna, ukuran dan rasa.

Bentuk, warna, ukuran dan rasa adalah besaran nominal, yaitu bersifat kategoris dan tiap nilai tidak bisa dijumlahkan atau dikurangkan. Dalam atribut warna ada beberapa nilai yang mungkin yaitu hijau, kuning, merah. Dalam atribut ukuran ada nilai besar, sedang dan kecil. Dengan nilai-nilai atribut ini, kemudian dibuat decision tree untuk menentukan suatu obyek termasuk jenis buah apa jika nilai tiap-tiap atribut diberikan (Santosa, 2007)

#### **D. Algoritma ID3**

Algoritma ID3 merupakan algoritma yang dipergunakan untuk membangun sebuah decision tree atau pohon keputusan. Algoritma ini ditemukan oleh J. Ross Quinlan (Widiyati, Wati, & Pakpahan, 2018), dengan memanfaatkan Teori

Informasi atau Information Theory milik Shanon. ID3 sendiri merupakan singkatan dari Iterative Dichotomiser 3.

Decision tree menggunakan struktur hierarki untuk pembelajaran supervised. Proses dari decision tree dimulai dari root node hingga leaf node yang dilakukan secara rekursif. Di mana setiap percabangan menyatakan suatu kondisi yang harus dipenuhi dan pada setiap ujung pohon menyatakan kelas dari suatu data.

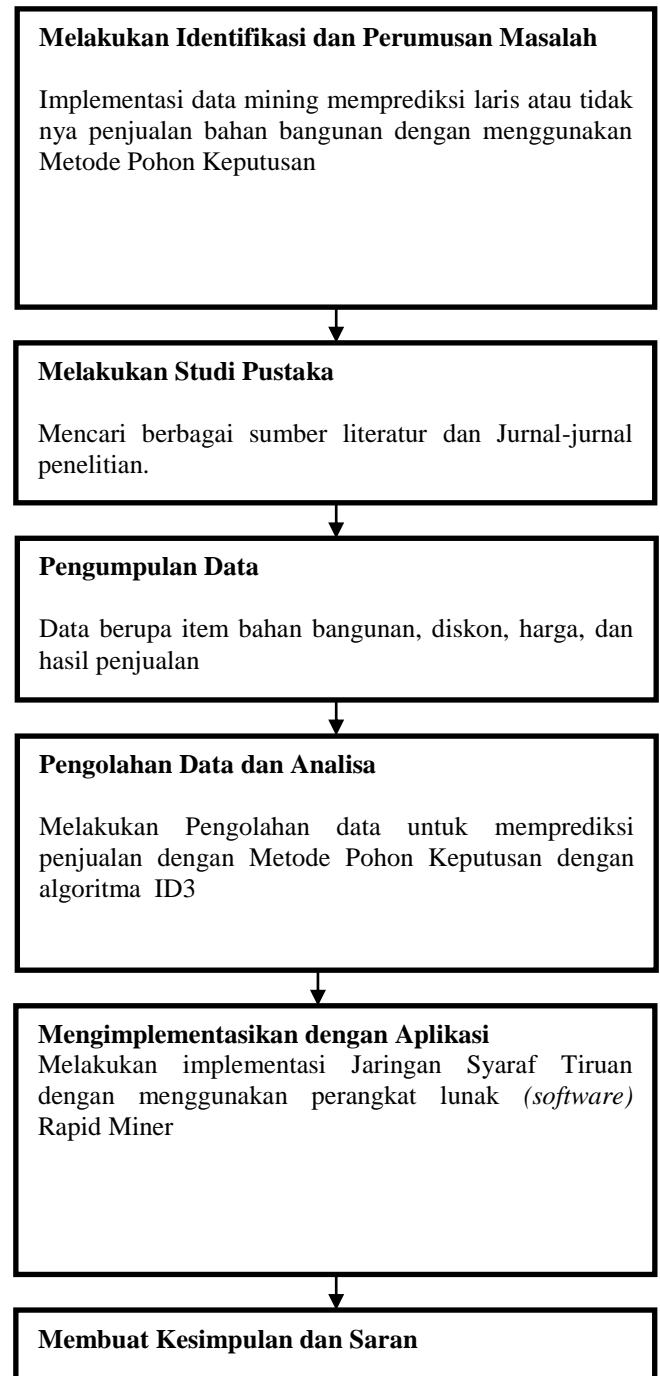
Proses dalam decision tree yaitu mengubah bentuk data (tabel) menjadi model pohon (tree) kemudian mengubah model pohon tersebut menjadi aturan (rule).

Dengan pendekatan ini, salah satu kelemahan algoritma dari decision tree, adalah faktor skalabilitas dimana algoritma tersebut hanya dapat digunakan untuk menangani sampel-sampel yang dapat disimpan secara keseluruhan dan pada waktu yang bersamaan di memori.

## 2. ISI PENELITIAN

### A. Kerangka Penelitian

Agar penelitian ini terarah, maka diperlukan kerangka kerja penelitian seperti dibawah ini :



Gambar 3.1 Kerangka Penelitian

### B. Tahapan Penelitian

#### 1. Pengumpulan Data

Metode yang digunakan dalam pengumpulan data adalah mencari dan menggunakan sumber data primer dan sumber data sekunder.

##### a. Sumber Data Primer

Data primer adalah data yang diperoleh langsung dari sumber melalui pengamatan, pencatatan

maupun dengan wawancara terhadap objek penulisan data tersebut diperoleh dengan teknik sebagai berikut:

1) Wawancara

Melakukan Wawancara langsung kepada pihak pekerja CV. Mitra Baja Cemerlang.

2) Survei

Melakukan survei lapangan dengan melihat penjualan obat-obatan di CV. Mitra Baja Cemerlang.

b. Sumber Data Sekunder

Selain sumber data primer, penyusunan penelitian ini juga mencari dan mengumpulkan data sekunder.

1) Penelitian lapangan ( *Field Research* )

Tahap awal yang dilakukan adalah observasi lapangan dengan mengadakan pengamatan langsung sesuai dengan kondisi saat melakukan observasi.

2) Penelitian Pustaka ( *Library Reasearch* )

Penelitian ini yang dilakukan melalui literatur-literatur yang berhubungan dengan tema penelitian untuk mencari informasi menyusun teori-teori yang berhubungan dengan pembahasan sehingga terjadi perpaduan ( interaksi ) yang kompleks antara satu dengan lainnya.

3) Penelitian Laboratorium ( *Laboratory Reasearch* )

Penelitian ini dilakukan dengan melakukan instalisasi editor untuk merancang perangkat lunak, dan melakukan instalasi database yang digunakan untuk membuat aplikasi yang dirancang.

2. Analisa dan Pengolahan Data

Pada pembuatan aplikasi ini analisa yang diperlukan adalah analisa *decision tree*. Karena pembuatan *data mining* ini diproses menggunakan metode *decision tree*. *Decision tree* memiliki dua jenis atribut yaitu : atribut *input* dan atribut *output*. Atribut *input* merupakan atribut pendukung dari

atribut target dan tentunya mendukung masalah yang ada, fungsinya sebagai pembanding dalam perhitungan *Gain Ratio*. Dalam data *training set* atribut *sample* (harus unik) dan minimal harus memiliki satu atribut target yang nilainya merupakan kesimpulan sementara permasalahan dari setiap *instance (record)*, dalam penelitian ini nilai dari atribut target adalah: Nilai Laris yang bernilai Ya dan Tidak. Atribut *input* yang memiliki *Gain Ratio* yang terbesar adalah atribut yang menjadi splitter terbaik.

Data yang akan diolah dapat dilihat pada Tabel 1.

Tabel 1. Data Kasus

Bagian	Type	Produksi	Size	Laris
Rangka	Kanal	Taso	0.75	T
Rangka	Kanal	Taso	0.65	T
Rangka	Kanal	Prima	0.75	F
Rangka	Kanal	Star	0.75	F
Rangka	Kanal	Kaso	0.75	T
Rangka	Kanal	Kaso	0.65	T
Rangka	Reng	Taso	0.45	T
Rangka	Reng	Prima	0.45	F
Rangka	Reng	Prima	0.40	T
Rangka	Reng	Kaso	0.45	T
Atap	Deck	Star	0.25	F
Atap	Deck	Star	0.30	F
Atap	Deck	Star	0.35	T
Atap	Deck	Star	0.40	T
Atap	Deck	Prima	0.25	T
Atap	Deck	Prima	0.30	F
Atap	Deck	Prima	0.35	T
Atap	Deck	Prima	0.40	T
Atap	Deck	Metal	0.30	F
Atap	Deck	Metal	0.35	T
Plafon	Gypsum	Jayaboard	0.9	T
Plafon	Gypsum	Indoboard	0.9	T
Plafon	Gypsum	Indoboard	0.75	F

### 3. Pengolahan Data dengan Metode ID3

#### a. Menentukan Node Akar

Node akar ditentukan dengan cara menghitung informasi gain dan entropy dari seluruh atribut data untuk menentukan root/akar awal dari tree.

Rumus :

$$Gain(A) = Entropi(S) - \sum_{i=1}^k \frac{|S_i|}{|S|} \times Entropi(S_i)$$

Entropi dihitung dengan rumus :

$$Entropi(S) = - \sum_{j=1}^k p_j \log_2 p_j$$

Entropi (Total) di dapat dari rumus perhitungan :

$$= -(15/22) \times \log_2 (15/22) + -(7/22) \times \log_2 (7/22) \\ = 0,90239328$$

Tabel 2. Entropi Total

Total Kasus	Sum (ya)	Sum (tidak)	Entropi Total
22	15	7	0,90239328

Kemudian dilakukan analisis pada setiap atribut dan nilai-nilainya dan beserta menghitung nilai entropi dan gain nya :

Tabel 3. Entropi dan Gain untuk setiap atribut

Atribut	Nilai	Sum (Nilai)	Sum (T)	Sum (F)	Entropi Total	Gain
Total Kasus		22	15	7	0,902	
Bagian	Rangka	10	7	3	0,881	0,060
	Atap	10	6	4	0,971	
	Plafon	2	2	0	0	
Type	Kanal	6	4	2	0,918	0,314
	Reng	4	3	1	0,811	
	Deck	10	6	4	0,971	
	Gypsum	2	2	0	0	
Produksi	Taso	3	3	0	0	0,277
	Prima	7	4	3	0,985	
	Star	5	2	3	0,971	
	Kaso	3	3	0	0	
	Metal	2	1	1	1	
	Jayaboard	1	1	0	0	
	Indoboard	1	1	0	0	
Ukuran	0.25	2	1	1	1	0,504
	0.30	3	0	3	0	
	0.35	3	3	0	0	
	0.40	3	3	0	0	
	0.45	3	2	1	0,918	
	0.65	2	2	0	0	
	0.75	4	2	2	1	
	0.9	2	2	0	0	

- Gain (Bagian) = 0,902 - ((10/22) x 0,881 + (10/22) x 0,971 + (2/22) x 0)) = 0,060
- Gain (Type) = 0,902 - ((6/22) x 0,918 + (4/22) x 0,811 + (10/22) x 0,971 + (2/22) x 0)) = 0,060
- Gain (Produksi) = 0,902 - ((3/22) x 0 + (7/22) x 0,985 + (5/22) x 0,971) + (3/22) x 0 + (2/22) x 1 + (1/22) x 0 + (2/22) x 0)) = 0,277
- Gain (Ukuran) = 0,902 - ((2/22) x 1 + (3/22) x 0 + (3/22) x 0 + (3/22) x 0 + (3/22) x 0,918 + (2/22) x 0 + (4/22) x 1 + (2/22) x 0)) = 0,504

Dari hasil perhitungan Gain di atas, dapat diketahui bahwa atribut dengan Gain tertinggi

adalah Gain (Ukuran), yaitu sebesar 0.504. Dengan demikian, Ukuran menjadi node akar.

Tabel 4. Atribut Ukuran untuk setiap nilai

Ukuran	0.25	2	1	1
	0.3	3	0	3
	0.35	3	3	0
	0.4	3	3	0
	0.45	3	2	1
	0.65	2	2	0
	0.75	4	2	2
	0.9	2	2	0

Dari tabel 3 dapat dilihat bahwa gain Ukuran memiliki beberapa atribut, yaitu :

- Atribut 0.30 semuanya bernilai tidak,
- Atribut 0.35 semuanya bernilai ya,
- Atribut 0.40 semuanya bernilai ya,
- Atribut 0.65 semuanya bernilai ya,
- Atribut 0.9 semuanya bernilai ya,

sehingga tidak perlu dilakukan perhitungan lebih lanjut.

Dengan demikian “Ukuran” 0.30, 0.35, 0.40, 0.65, 0.9 menjadi daun atau *leaf*.

Sedangkan nilai atribut 0.25, 0.45, 0.9 belum mengklasifikasikan kasus menjadi satu keputusan sehingga perlu dilakukan perhitungan lagi.

#### b. Mencari Node Cabang

Untuk selanjutnya dilakukan pencarian node cabang yang dilakukan untuk mencari node cabang dari nilai atribut Ukuran. Perhitungan dilakukan dengan mencari nilai dari atribut selain “Ukuran” 0.30, 0.35, 0.40, 0.65, 0.9.

Sebagai contoh hasil perhitungan dari node cabang Ukuran 0.45, sebagai berikut :

Tabel 5. Atribut Ukuran untuk nilai 0.45

Bagian	Type	Produksi	Size	Laris
Rangka	Reng	Taso	0.45	T
Rangka	Reng	Prima	0.45	F
Rangka	Reng	Kaso	0.45	T

Tabel 6. Entropi dan Gain Ukuran 0.45

Atribut	Nilai	Sum Nilai	Sum (T)	Sum (F)	Entropi Total	Gain
Total Kasus		3	2	1	0,918	
Bagian Type	Rangka	3	2	1	0,918	0,793
	Reng	3	2	1	0,918	
Produksi	Taso	1	1	0	0	0,918
	Prima	1	0	1	0	
	Kaso	1	1	0	0	

### 3. KESIMPULAN

Dari uraian penelitian di atas dapat ditarik kesimpulan sebagai berikut :

1. Pengolahan data penjualan barang dengan data mining metode decision tree menggunakan algoritma ID3 dapat digunakan untuk memprediksi penjualan bahan bangunan baja ringan yang laris atau tidak pada CV. Mitra Baja Cemerlang.
2. Implementasi data mining ini menjadi media efektif untuk mengembangkan perusahaan dan meningkatkan pelayanan kepada pelanggan karena implementasi Data Mining ini dapat menyajikan data secara cepat dan memberikan informasi bagi perusahaan dengan proses perhitungan sesuai gain yang tertinggi dipilih sebagai root pada hasil pohon keputusan yang dibuat dan pada akhirnya pemilik perusahaan bisa mengambil keputusan.

Implementasi Data Mining ini dapat digunakan sebagai penunjang keputusan pada pemilik perusahaan dalam mengambil keputusan untuk menentukan persediaan baja ringan pada bulan berikutnya berdasarkan data-data penjualan barang yang laris

### DAFTAR PUSTAKA

- [1] A. Rufaidah and A. Fatakh, “Analisis Pengendalian Persediaan Bahan Baku Dengan Menggunakan Metode Economic Order Quantity (EOQ) Di PT. X,” KAIZEN Manag. Syst. Ind. Eng. J., vol. 1, no. 2, pp. 40–45, 2018.

- [2] N. Apriyani and A. Muhsin, "Analisis Pengendalian Persediaan Bahan Baku Dengan Metode Economic Order Quantity Dan Kanban Pada PT. Adyawinsa Stamping Industries," *J. Optimasi Sist. Ind.*, vol. 10, no. 2, pp. 128–142, 2017.
- [3] D. R. Indah, L. Purwasih, and Z. Maulida, "Pengendalian Persediaan Bahan Baku Pada PT. Aceh Rubber Industries Kabupaten Aceh Tamiang," *J. Manaj. dan Keuang.*, vol. 7, no. 2, pp. 157–173, 2018.
- [4] D. K. Widiyati, M. Wati, and H. S. Pakpahan, "Penerapan Algoritma ID3 Decision Tree pada Penentuan Penerimaan Program Bantuan Pemerintah Daerah di Kabupaten Kutai Kartanegara," *JURTI*, vol. 2, no. 2, pp. 125–134, 2018.
- [5] M. S. Efendi and H. A. Wibawa, "Prediksi Penyakit Diabetes Menggunakan Algoritma ID3 dengan Pemilihan Atribut Terbaik (Diabetes Prediction using ID3 Algorithm with Best Attribute Selection)," *Juita*, vol. VI, no. 1, pp. 29–35, 2018.
- [6] I. R. Munthe and V. Sihombing, "Klasifikasi Algoritma Iterative Dichotomizer ( ID3 ) untuk Tingkat kepuasan pada Sarana Laboratorium Komputer," *Jutikomp*, vol. 1, no. 2, pp. 27–34, 2018.
- [7] Wahyudin, "Metode Iterative Dichotomizer 3 ( ID3) Untuk Penyeleksian Penerimaan Mahasiswa Baru," *J. Pendidik. Teknol. Inf. dan Komun.*, vol. 2, no. 2, pp. 5–15, 2009.
- [8] W. W. Ariestya, Y. E. Praptiningsih, and W. Supriatin, "Decision Tree Learning Untuk Penentuan Jalur Kelulusan Mahasiswa," *J. Ilm. FIFO*, vol. 8, no. 1, pp. 14–21, 2016.
- [9] T. D. Utama, S. W. Sihwi, and A. Doewes, "Implementasi Algoritma Iterative Dichotomiser 3 Pada Penyeleksian Program Mahasiswa Wirausaha UNS," *J. ITSMART*, vol. 3, no. 2, pp. 74–82, 2014.
- [10] E. Indra, K. Ho, Arlinanda, R. Hakim, D. Sitanggang, and O. Sihombing, "Application of C4.5 Algorithm for Cattle Disease Classification," *J. Phys. Conf. Ser.*, vol. 1230, p. 012070, Jul. 2019, doi: 10.1088/1742-6596/1230/1/012070.
- [11] E. Indra and Y. Laia, "Aplikasi Untuk Mengetahui Mahasiswa Berpotensi Drop Out Dengan Metode K-Means Pada Program Studi Sistem Informasi Universitas Prima Indonesia," *J. Senopati Semin. Nas. Pascasarj. Tek. Inform.*, vol. 1, no. 1, p. 6, 2016.